



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(3) (2026)



ijerst.editor@gmail.com
editor@ijerst.com

Research Paper

A Context-Driven Artificial Intelligence Framework for Detecting Fraudulent Activities in Payment Systems

S. Venkata Achuta Rao^{1*}, P Sravan Kumar², Guttula Shruthula², Kummari Ramakrishna²,
Maheswarapu Sowmya², Islavath Siddu²

¹Professor and Dean of Academics, Department of Computer Science and Engineering (AI&ML),
Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana,
India

²Department of Electronics and Communication Engineering, Sree Dattha Institute of Engineering
and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana, India

*Correspondence: S. Venkata Achuta Rao (sreedatthaachyuth@gmail.com)

ABSTRACT

The expansion of digital payment platforms has created a growing need for robust and reliable fraud detection systems to safeguard financial transactions. Earlier approaches to fraud detection were primarily based on manual verification, rule-driven mechanisms, and simple statistical models. These methods required substantial human involvement, were time-intensive, and struggled to identify sophisticated fraud patterns within large and continuously evolving transaction datasets. As the volume and complexity of digital payments increased, these traditional techniques became inadequate, particularly in supporting real-time decision-making. To address these limitations, this work proposes an intelligent web-based framework for comprehensive fraud detection in Unified Payments Interface (UPI) transactions by leveraging advanced machine learning and deep learning methodologies. The system is implemented using the Flask framework and incorporates natural language processing techniques to process textual transaction data. For feature representation, Sentence Bidirectional Encoder Representations from Transformers (SBERT) is employed to generate contextual embeddings that capture semantic relationships within the data. These embeddings are then used to train and evaluate multiple classification models, including Gaussian Naive Bayes (GNB), Bernoulli Naive Bayes (BNB), Multinomial Naive Bayes (MNB), and a Histogram-based Gradient Boosting (HGB) classifier. The framework supports both binary classification for fraud identification and multi-class classification for categorizing transaction types. Experimental results demonstrate that the HGB model achieves superior performance, delivering high accuracy along with strong Precision, Recall, and F1-score compared to conventional approaches. Additionally, the system provides a secure web interface that allows users to upload datasets and obtain real-time predictions, with options to download results for further analysis. Overall, the proposed solution enhances automation, scalability, and predictive accuracy, contributing to more effective and efficient fraud detection in digital financial systems.

Keywords: Unified Payments Interface (UPI), Digital Payments, Machine Learning (ML), Natural Language Processing (NLP), Histogram-Based Gradient Boosting (HGB).

1. INTRODUCTION

The continuous evolution of technology, along with the widespread use of digital payment platforms, has significantly changed the way financial transactions are carried out. While these advancements have improved convenience and accessibility, they have also contributed to a noticeable increase in digital payment fraud. Cyber attackers are constantly refining their methods to exploit weaknesses in payment infrastructures, creating serious concerns regarding the safety and dependability of online financial systems. This study examines different strategies for detecting and preventing fraud in digital payment environments, particularly within the context of the modern digital age and Industry 4.0.

The main goal of this research is to gain a deeper understanding of the risks affecting digital payment systems and to develop effective mechanisms to reduce these threats. A critical step in this process involves identifying common forms of fraudulent activities associated with digital transactions. These include identity theft, account hijacking, phishing schemes, and malware-based attacks. Such threats have a substantial impact on individuals, organizations, and financial institutions, often resulting in monetary losses, damage to reputation, and reduced customer confidence.

With the rise of the fourth industrial revolution, advanced technologies such as Artificial Intelligence and Big Data Analytics have reshaped the landscape of fraud detection. These technologies enable the processing of large-scale data and provide powerful analytical tools to detect irregular patterns, anomalies, and suspicious activities in a timely manner. However, their adoption also brings additional complexity and introduces new security challenges. As a result, modern fraud detection systems must be designed to be intelligent, scalable, and resilient in order to effectively address evolving threats and maintain secure digital financial operations.

2. Related Work

The rapid growth of digital financial transactions has significantly increased the risk of fraudulent activities, necessitating the adoption of intelligent and adaptive fraud detection systems. Earlier approaches primarily relied on rule-based and statistical methods; however, these techniques lack the flexibility to adapt to evolving fraud patterns. Recent advancements in machine learning (ML) and deep learning (DL) have enabled the development of more robust, scalable, and real-time fraud detection frameworks.

2.1 Machine Learning-Based Fraud Detection

Traditional ML models have been widely applied for fraud detection due to their simplicity and effectiveness. Binsawad *et al.* proposed the Voted Perceptron (VP) model, which dynamically updates decision boundaries based on misclassified samples, resulting in improved classification performance compared to models such as A1DE, KNN, NB, RT, and FT [1]. Similarly, Abd Razak *et al.* conducted a systematic literature review and identified Support Vector Machine (SVM) and Artificial Neural Networks (ANN) as the most commonly used techniques in fraud detection systems [7].

In addition, Golightly *et al.* emphasized the importance of balancing evaluation metrics such as accuracy, recall, and F1-score, particularly in cybersecurity contexts where class imbalance is a major challenge [14]. Chang *et al.* further addressed this issue by comparing random under-sampling and SMOTE, demonstrating that SMOTE provides a better balance between precision and recall, while under-sampling improves detection sensitivity [8].

2.2 Deep Learning and Hybrid Models

To overcome the limitations of traditional ML approaches, several studies have focused on DL and hybrid architectures. Abbassi *et al.* introduced a hybrid VAE-QLSTM model that combines representation learning with temporal sequence modeling, achieving high accuracy and sensitivity on real-world datasets [2], [9], [15]. Similarly, Sathupadi *et al.* developed BankNet, which integrates big data processing with BiLSTM networks to achieve high detection accuracy and low latency in large-scale financial systems [6].

Chu *et al.* proposed a multi-layered intelligent detection framework incorporating feature selection and ensemble voting mechanisms, significantly improving classification performance in noisy financial environments [10]. These hybrid models demonstrate the effectiveness of combining multiple learning paradigms to enhance fraud detection capabilities.

2.3 Knowledge-Driven and Adaptive Approaches

In addition to data-driven methods, incorporating domain knowledge has shown promising results. Sun *et al.* enhanced boosting algorithms by integrating prior human knowledge, such as expert rules and historical fraud patterns, leading to notable improvements in accuracy and recall [5]. Furthermore,

Oztemel *et al.* highlighted the importance of dynamic and agent-based models, which can continuously adapt to changing fraud behaviors, unlike static analytical systems [3].

These approaches emphasize the need for adaptive learning mechanisms that can evolve alongside emerging fraud strategies.

2.4 Privacy, Security, and Ethical Considerations

With the increasing use of AI in financial systems, privacy and ethical concerns have become critical. Yaseen *et al.* stressed the importance of explainable and fairness-aware AI models to ensure transparency and regulatory compliance [4]. Abbassi *et al.* and AbouGrad *et al.* explored decentralized and federated learning frameworks, enabling collaborative fraud detection without sharing sensitive data across institutions [9], [12].

Additionally, Theodorakopoulos *et al.* discussed challenges such as overfitting, limited data availability, and real-time deployment, recommending distributed and ensemble-based approaches to improve system robustness [11].

2.5 Research Gap

Although significant progress has been made in fraud detection using ML and DL techniques, several limitations remain. Existing systems often struggle with class imbalance, high computational complexity, and lack of real-time adaptability. Moreover, many advanced models rely heavily on centralized architectures, raising concerns about data privacy and scalability. While hybrid and federated learning approaches address some of these challenges, there is still a need for a lightweight, efficient, and scalable framework that balances detection accuracy, computational cost, and real-time performance.

The proposed system aims to address these limitations by integrating optimized ML models with efficient preprocessing and evaluation strategies to achieve improved fraud detection performance with reduced complexity.

3. PROPOSED SYSTEM

The proposed approach presents a hybrid multi-output classification framework designed to jointly predict both transaction type and fraud status within a unified processing pipeline, as illustrated in Fig. 1. In contrast to traditional rule-based systems that depend on fixed rules and primarily utilize structured numerical data, and unlike conventional machine learning methods that typically address a single prediction task, this framework integrates semantic text representations with structured feature analysis. Specifically, contextual embeddings generated using SBERT are combined with numerical transaction features to enhance the overall representation of the data. To address the challenge of class imbalance, the system incorporates SMOTE, which improves the learning capability of the models. A range of classification algorithms, including GNB, BNB, MNB, and HGB, are employed for prediction, with HGB demonstrating the best performance. This integrated methodology supports both binary and multi-class classification simultaneously, leverages both structured and unstructured data sources, adapts to emerging fraud patterns without requiring manual rule updates, and significantly improves prediction accuracy and reliability compared to existing approaches.

Data Acquisition and Representation: The system begins by collecting transaction data containing both structured attributes (amount, age groups, timestamps, merchant identifiers) and unstructured attributes (transaction descriptions, remarks). All features are preserved to avoid information loss, ensuring that both numeric trends and linguistic cues are available for learning.

Preprocessing and Data Normalization: Structured fields are cleaned by handling missing values according to their type, while age group information is decomposed into numerical lower and upper bounds to retain quantitative meaning. Unstructured text undergoes a cleaning process that includes normalization, removal of irrelevant characters, elimination of stopwords, and lemmatization to reduce linguistic variability.

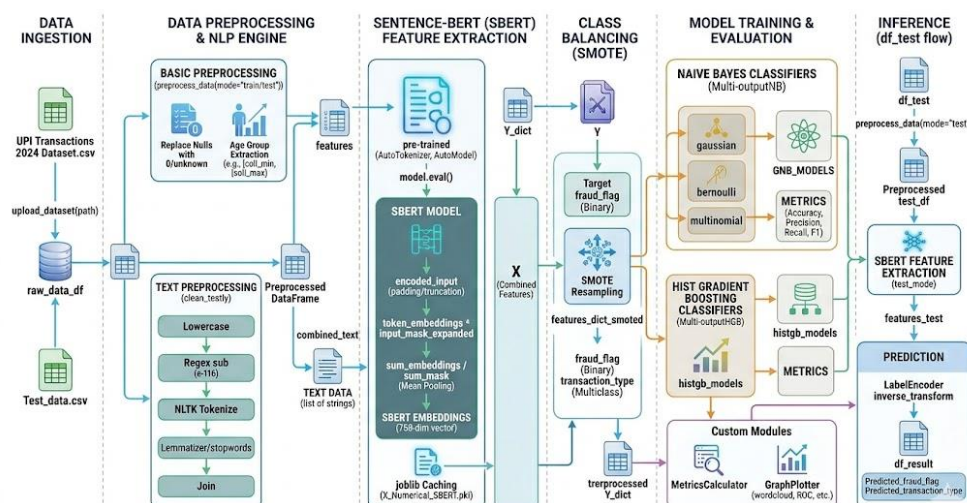


Fig. 1: Proposed System Architecture

Semantic Feature Extraction: The processed text is transformed into dense semantic vectors using Numerical SBERT. This step captures deep contextual meaning, enabling the system to detect subtle fraud indicators and transaction type patterns that static keyword-based or rule-driven approaches miss. Unlike older embedding methods, Numerical SBERT generates context-sensitive representations that can differentiate between superficially similar but semantically distinct transaction notes.

Feature Fusion: The semantic vectors from Numerical SBERT are concatenated with the cleaned numerical features, producing a unified multi-modal feature space. This fusion allows the model to reason jointly over numeric trends (e.g., unusually high amounts) and textual semantics (e.g., suspicious note descriptions), creating a richer decision-making foundation.

Imbalance Handling with SMOTE: Since fraudulent transactions are rare compared to normal ones, the combined dataset is balanced using SMOTE. This oversampling method synthetically generates realistic samples of minority classes in both prediction targets, preventing the classifier from becoming biased toward majority classes and improving its ability to detect rare but critical fraud cases.

Multi-Output Model Training: Two related classification tasks fraud detection and transaction type classification are trained concurrently using the same feature set. By treating them as related but separate outputs, the system benefits from shared feature learning while still tailoring decision boundaries for each task. Models such as ensemble gradient boosting classifiers are applied, which can handle complex non-linear relationships across both numeric and semantic inputs.

Evaluation and Metrics Analysis: Model predictions are evaluated using comprehensive metrics including accuracy, precision, recall, and F1-score for each output. This ensures that performance is measured not only by overall correctness but also by its ability to correctly identify minority classes, directly addressing the precision recall trade-off often ignored in traditional systems.

Deployment and Adaptability: The trained multi-output model can be deployed to process incoming UPI transactions in real time, producing both fraud risk labels and transaction category predictions. Unlike traditional rule-based systems, it adapts automatically to new patterns as long as it is retrained with updated data, eliminating the need for frequent manual rule rewriting.

HGB Classifier

The HGB Classifier is an ensemble learning algorithm that builds a series of decision trees sequentially, where each tree tries to correct the errors of the previous ones. Its internal operation leverages histogram-based binning of continuous features to speed up training and reduce memory usage. The model iteratively optimizes a loss function using gradient descent principles, combining weak learners into a strong predictive model.

Step-1: Feature Binning

The model starts by discretizing continuous input features into bins (histograms). This reduces the number of unique feature values, allowing the algorithm to compute split points efficiently and accelerate tree construction.

Step-2: Initial Prediction

An initial prediction is generated for all training samples, typically using a simple constant (e.g., the average target value in regression or the prior probability in classification). This provides a baseline from which errors (residuals) are computed.

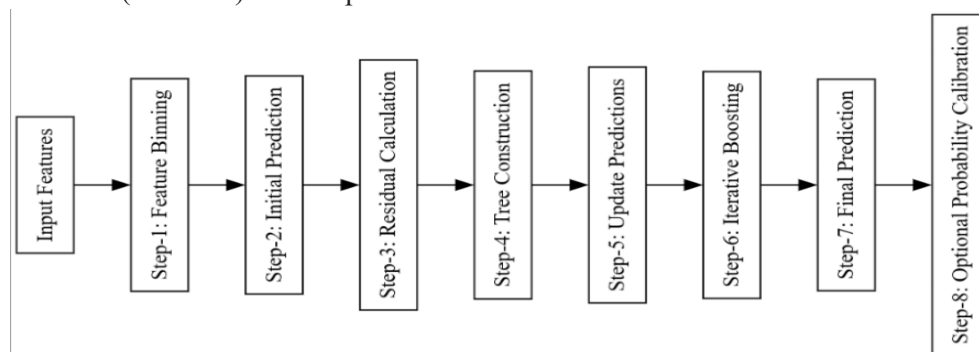


Fig. 2: Internal operational workflow of HGB Classifier.

Step-3: Residual Calculation

For each training sample, the model calculates the pseudo-residuals, which are the gradients of the loss function with respect to the current prediction. These residuals quantify how far off the current predictions are from the true target values.

Step-4: Tree Construction

A decision tree is constructed to predict these residuals. The tree selects splits that minimize the loss function, using the binned feature representation to speed up the search for optimal splits. This tree acts as a weak learner in the ensemble.

Step-5: Update Predictions

The predictions from the newly grown tree are scaled by a learning rate and added to the existing predictions. This step incrementally improves the overall model, moving predictions closer to the actual targets.

Step-6: Iterative Boosting

Steps 3–5 are repeated for a predefined number of iterations (or until convergence). Each new tree focuses on correcting the errors of the combined previous trees, progressively refining the model.

Step-7: Final Prediction

After all boosting iterations, the ensemble of trees produces the final prediction. For classification, the outputs from all trees are combined (often using a softmax function) to generate class probabilities and the final class label.

Results description

Fig. 3 shows two count plots representing the distribution of target columns: (a) Fraud flag and (b) Transaction type. In the Fraud flag plot, there is a significant imbalance, with a count of approximately 10,000 for class 0 (non-fraud) and a negligible count for class 1 (fraud), indicating a highly skewed dataset where non-fraud cases dominate. The Transaction type plot (b) displays a more balanced distribution across four classes, with class 2 having the highest count around 5,000, followed by class 1 at about 4,000, class 0 at around 2,000, and class 3 with the lowest count near 1,000, suggesting varying frequencies of different transaction types in the dataset.

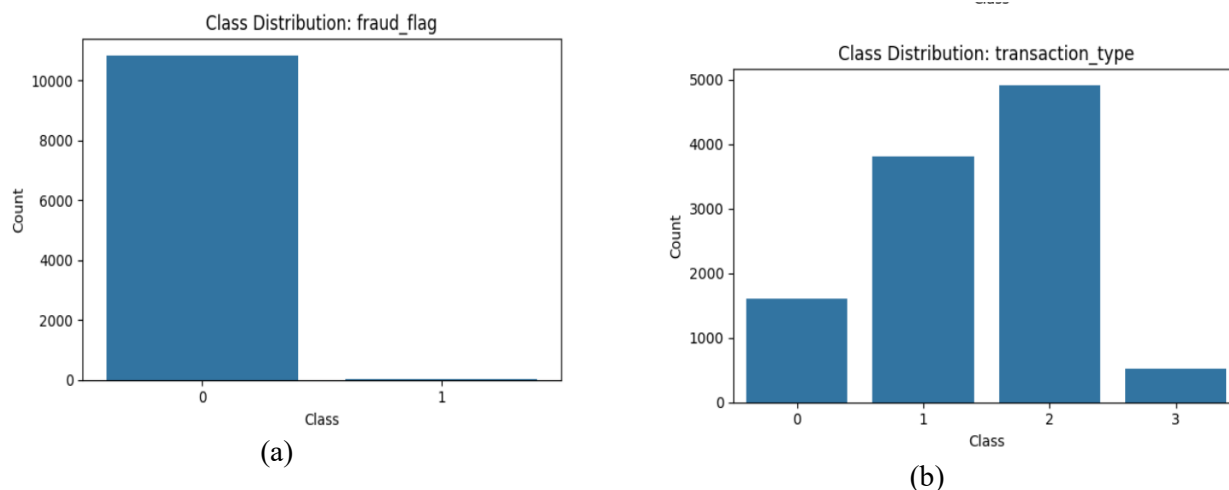


Fig. 3: Count plot of target columns (a) Fraud flag. (b) Transaction type.

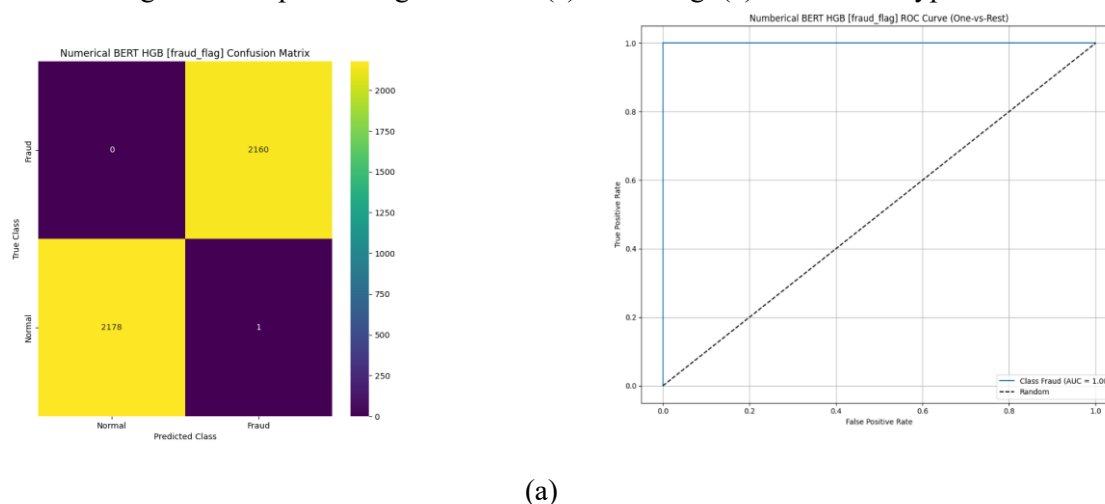


Figure 4 (a) - Numerical BERT HGB [fraud_flag] Confusion Matrix and ROC Curve (One-vs-Rest): The confusion matrix details show Fraud with 0 true positives and 2160 false positives, and Normal with 2178 true negatives and 1 false negative for predicted classes Normal and Fraud respectively. The ROC curve indicates an AUC of 1.00 for the fraud class, far exceeding the random classifier (AUC = 0.5), demonstrating perfect discrimination between fraud and normal transactions.

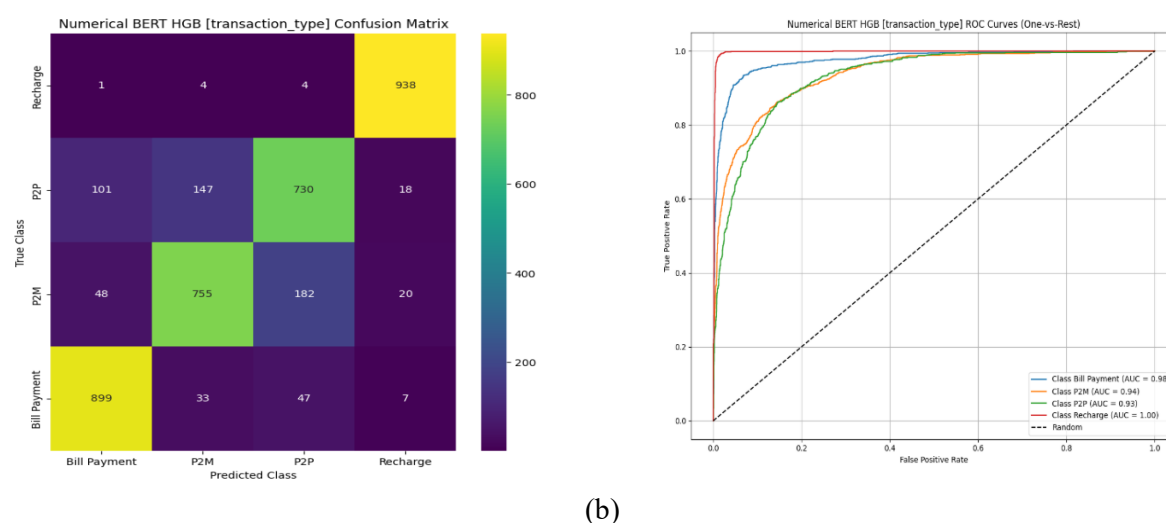


Fig. 4: Confusion matrices and AUC-ROC curves obtained using proposed Multi-output HGB Classifier for both target columns. (a) Fraud flag. (b) Transaction type.

Figure 4 (b) - Numerical BERT HGB [transaction_type] Confusion Matrix and ROC Curves (One-vs-Rest): The confusion matrix details show Bill Payment with 899 correct predictions, 33 misclassified as P2M, 47 as P2P, and 7 as Recharge; P2M with 48 correct, 755 as Bill Payment, 182 as P2P, and 20 as Recharge; P2P with 101 correct, 147 as Bill Payment, 730 as P2M, and 18 as Recharge; and Recharge with 1 correct, 4 as Bill Payment, 4 as P2M, 4 as P2P, and 938 as Recharge for predicted classes Bill Payment, P2M, P2P, and Recharge respectively. The ROC curves indicate AUC values of 0.98 for Bill Payment, 0.94 for P2M, 0.91 for P2P, and 1.00 for Recharge, all significantly above the random classifier (AUC = 0.5), showcasing excellent discriminative power across transaction types.

Fig. 5 shows the output interface displaying the fraud detection results in tabular form. The table includes features like HOUR_OF_DAY, DAY_OF_WEEK, IS_WEEKEND, PREDICTED_TRANSACTION_TYPE, and PREDICTED_FRAUD_FLAG. Each record reflects the model's classification of transactions as normal or fraudulent. This page helps users easily interpret and verify prediction outcomes. It demonstrates the final stage of the analytical process and the model's ability to detect fraudulent behavior accurately.

Prediction Results				
HOUR_OF_DAY	DAY_OF_WEEK	IS_WEEKEND	PREDICTED_TRANSACTION_TYPE	PREDICTED_FRAUD_FLAG
20	Saturday	1	P2P	Normal
9	Monday	0	P2P	Normal
17	Thursday	0	P2P	Normal
10	Thursday	0	P2P	Normal
10	Sunday	1	Bill Payment	Normal
13	Thursday	0	P2P	Normal
0	Friday	0	Bill Payment	Normal
13	Friday	0	Bill Payment	Normal
9	Monday	0	P2P	Normal

Fig. 5: Predictions on test data.

The performance comparison in Table 1 highlights the effectiveness of various models for the 'fraud flag' target column, with the Numerical BERT Hist Gradient Boosting (HGB) model achieving the highest scores across all metrics, boasting an accuracy, precision, recall, and F1-score of 99.977%, indicating near-perfect classification performance. The Numerical BERT GNB model follows with a solid performance, recording an accuracy of 85.481%, precision of 85.838%, recall of 85.459%, and an F1-score of 85.439%, suggesting a balanced and reliable detection capability for fraud. In contrast, the Numerical BERT BNB model underperforms significantly, with an accuracy of 49.781%, precision of 24.891%, recall of 50.000%, and an F1-score of 33.236%, reflecting poor discriminative power and likely random-like predictions. The Numerical BERT MNB model falls in the middle, with an accuracy of 75.686%, precision of 76.330%, recall of 75.651%, and an F1-score of 75.519%, indicating moderate performance but still lagging behind the Gaussian NB and HGB models. Overall, the HGB model stands out as the most effective, while the MNB model appears unsuitable for this task, with GNB and MNB offering intermediate performance levels.

Table. 1: Performance comparison of existing and proposed models for fraud flag target column.

Algorithm	Accuracy	Precision	Recall	F1-Score
Numerical BERT GNB [fraud flag]	85.481	85.838	85.459	85.439
Numerical BERT BNB [fraud flag]	49.781	24.891	50.000	33.236
Numerical BERT MNB [fraud flag]	75.686	76.330	75.651	75.519
Numerical BERT HGB [fraud flag]	99.977	99.977	99.977	99.977

Table. 2: Performance comparison of existing and proposed models for transaction type target column.

Algorithm	Accuracy	Precision	Recall	F1-Score
Numerical BERT GNB [transaction type]	30.325	30.275	30.587	29.337
Numerical BERT BNB [transaction type]	24.072	6.018	25.000	9.701
Numerical BERT MNB [transaction type]	28.317	28.547	28.544	27.431
Numerical BERT HGB [transaction type]	84.443	84.333	84.661	84.440

The performance comparison in Table 2 evaluates the effectiveness of various models for the 'transaction type' target column, with the Numerical BERT HGB model leading with an accuracy of 84.443%, precision of 84.333%, recall of 84.661%, and an F1-score of 84.440%, demonstrating superior classification performance across multiple transaction categories. The Numerical BERT GNB model follows with a modest performance, achieving an accuracy of 30.325%, precision of 30.275%, recall of 30.587%, and an F1-score of 29.337%, indicating limited but balanced predictive capability. The Numerical BERT BNB model performs poorly, with an accuracy of 24.072%, precision of 6.018%, recall of 25.000%, and an F1-score of 9.701%, suggesting it struggles significantly with this multi-class classification task. The Numerical BERT MNB model also underperforms relative to HGB, with an accuracy of 28.317%, precision of 28.547%, recall of 28.544%, and an F1-score of 27.431%, showing only marginal improvement over GNB. Overall, the HGB model excels, while the other models, particularly BNB, exhibit substantial weaknesses in accurately classifying transaction types.

5. Conclusion

This study presents a reliable and efficient solution that combines natural language processing and machine learning techniques for multi-output classification of Unified Payments Interface transactions, addressing both fraud detection and transaction type identification. By applying advanced text preprocessing methods and leveraging contextual embeddings generated through SBERT, along with imbalance handling using SMOTE, the system achieves strong predictive performance across various models, including different Naive Bayes variants and the Histogram-Based Gradient Boosting classifier. The design follows a modular architecture, ensuring that each stage data preprocessing, feature extraction, model training, and evaluation remains independent, flexible, and reusable. Furthermore, the inclusion of both graphical and tabular evaluation outputs enhances interpretability, allowing analysts and stakeholders to better understand model effectiveness and performance metrics. Overall, the proposed framework provides a scalable, adaptable, and technically robust foundation for practical deployment in digital payment security systems, contributing to more accurate and efficient fraud detection in real-world scenarios.

REFERENCES

- [1] Binsawad M. Enhanced Financial Fraud Detection Using an Adaptive Voted Perceptron Model with Optimized Learning and Error Reduction. *Electronics*. 2025; 14(9):1875. <https://doi.org/10.3390/electronics14091875>
- [2] Abbassi H, El Mendili S, Gahi Y. Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion. *Big Data and Cognitive Computing*. 2025; 9(7):185. <https://doi.org/10.3390/bdcc9070185>

- [3] Oztemel E, Isik M. A Systematic Review of Intelligent Systems and Analytic Applications in Credit Card Fraud Detection. *Applied Sciences*. 2025; 15(3):1356. <https://doi.org/10.3390/app15031356>
- [4] Yaseen H, Al-Amarnah A. Adoption of Artificial Intelligence-Driven Fraud Detection in Banking: The Role of Trust, Transparency, and Fairness Perception in Financial Institutions in the United Arab Emirates and Qatar. *Journal of Risk and Financial Management*. 2025; 18(4):217. <https://doi.org/10.3390/jrfm18040217>
- [5] Sun Q, Tang T, Chai H, Wu J, Chen Y. Boosting Fraud Detection in Mobile Payment with Prior Knowledge. *Applied Sciences*. 2021; 11(10):4347. <https://doi.org/10.3390/app11104347>
- [6] Sathupadi K, Achar S, Bhaskaran SV, Faruqui N, Uddin J. BankNet: Real-Time Big Data Analytics for Secure Internet Banking. *Big Data and Cognitive Computing*. 2025; 9(2):24. <https://doi.org/10.3390/bdcc9020024>
- [7] Ali A, Abd Razak S, Othman SH, Eisa TAE, Al-Dhaqm A, Nasser M, Elhassan T, Elshafie H, Saif A. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Applied Sciences*. 2022; 12(19):9637. <https://doi.org/10.3390/app12199637>
- [8] Chang V, Ali B, Golightly L, Ganatra MA, Mohamed M. Investigating Credit Card Payment Fraud with Detection Methods Using Advanced Machine Learning. *Information*. 2024; 15(8):478. <https://doi.org/10.3390/info15080478>
- [9] Abbassi H, El Mendili S, Gahi Y. Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion. *Big Data and Cognitive Computing*. 2025; 9(7):185. <https://doi.org/10.3390/bdcc9070185>
- [10] Li X, Chu L, Li Y, Xing Z, Ding F, Li J, Ma B. An Intelligent Financial Fraud Detection Support System Based on Three-Level Relationship Penetration. *Mathematics*. 2024; 12(14):2195. <https://doi.org/10.3390/math12142195>
- [11] Theodorakopoulos L, Theodoropoulou A, Tsimakis A, Halkiopoulos C. Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost. *Electronics*. 2025; 14(9):1754. <https://doi.org/10.3390/electronics14091754>
- [12] AbouGrad H, Sankuru L. Online Banking Fraud Detection Model: Decentralized Machine Learning Framework to Enhance Effectiveness and Compliance with Data Privacy Regulations. *Mathematics*. 2025; 13(13):2110. <https://doi.org/10.3390/math13132110>
- [13] Aljunaid SK, Almheiri SJ, Dawood H, Khan MA. Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection. *Journal of Risk and Financial Management*. 2025; 18(4):179. <https://doi.org/10.3390/jrfm18040179>
- [14] Golightly L, Ganatra MA, Mohamed M. Investigating Credit Card Payment Fraud with Detection Methods Using Advanced Machine Learning. *Information*. 2024; 15(8):478. <https://doi.org/10.3390/info15080478>
- [15] Abbassi H, El Mendili S, Gahi Y. Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion. *Big Data and Cognitive Computing*. 2025; 9(7):185. <https://doi.org/10.3390/bdcc9070185>